

# Transforming in the Digital Tide: An In-depth Analysis of the Impact of E-Commerce on Brick-and-Mortar Retail in China, 2019-2023

Qingyue Chen 22-737-639

Shiyang Jin 22-737-712

Yuru Jin 21-749-650

Zhengyu Duan 22-736-904

December 2023

# Table of Contents

1 Background

2 Process

3 Analysis

- **Significant Change:** E-commerce reshapes China's retail landscape.
- **Focus:** Analyzing how e-commerce transforms traditional retail.
- **Analysis:** Examining retail trends, online/offline sales, consumer shifts.
- **Goal:** Detailed, data-driven insights into China's evolving retail sector.
- **China's E-commerce Boom:** From RMB 930 billion (2004) to RMB 29,160 billion (2017); 533 million online shoppers by 2017.
- **Global Dominance:** China's global e-commerce market share soared from under 1% to over 40%.

# Table of Contents

1 Background

2 Process

3 Analysis

## Assumption:

- Data Reliability, Consumer Behavior, Representation, Technology Access, Retail Segmentation.

## Predictions:

- Market Shift: Growing negative correlation between online and offline retail indicates a shift towards digital platforms.
- COVID-19 Impact: Pandemic accelerates e-commerce adoption, altering consumer behavior.

## Data Section Overview:

- **Data Sources:** Primary data from the National Bureau of Statistics of China.
- **Collection Method:** Efficient data gathering using Python-based web scraping tools.
- **Dataset Insights:** Interplay and trends between online and offline retail channels.

## Data Processing Overview:

- Integration and Summarization: Unifying various datasets.
- Handling Data Shortages: Averaging technique for January data.
- Correlation Analysis: Examining online-offline sales relationship.

## Data Crawling with Python:

- Data Extraction: HTTP POST requests for retail data.
- Parsing and Cleaning: Custom functions for JSON parsing.
- Data Structuring: Organizing data in pandas DataFrames.

## Key Considerations in Data Handling:

- Integration of Diverse Data Sources.
- Addressing Seasonal Data Shortages.
- Ensuring Cumulative Data Accuracy.
- Assessing COVID-19's Retail Impact.

# Table of Contents

1 Background

2 Process

3 Analysis

# Descriptive Analysis of Retail Sales

## Online Sales Trends:

- Monthly sales fluctuate with seasonal peaks.
- An increasing trend observed from 2019 to 2023.
- Data suggests potential seasonal patterns.



Figure: Monthly Online Sales Data



## In-Store Sales Patterns:

- Larger fluctuations in sales compared to online.
- Sharp increase at the start of 2020, hinting at a possible event impact or seasonal trend.
- Trends assist in comparing online vs. in-store performance and strategic planning.



Figure: Monthly Offline Sales Data

# Trends in In-Store vs. Online Retail Sales:

- Online sales are steadily increasing over time.
- In-store sales are more volatile and slightly declining.
- The gap between in-store and online sales is narrowing.

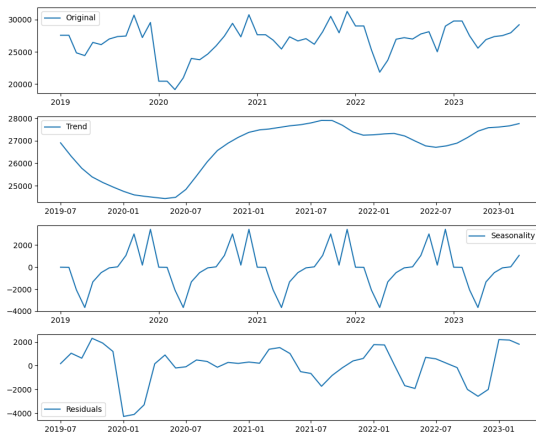


Figure: Online and offline sales compare

# Seasonal Analysis in Retail Sales

## Offline Trends:

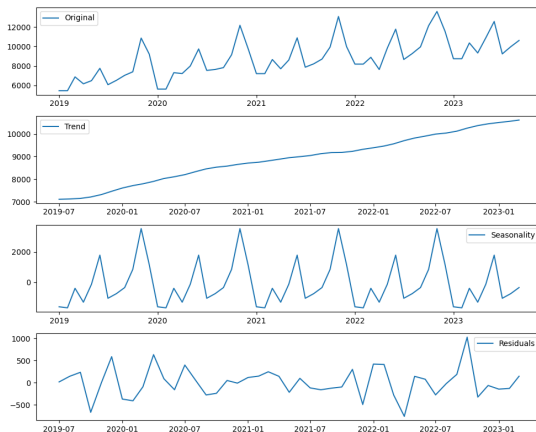
- Fluctuations in physical store sales with seasonal influences.
- Decline in early 2020 due to COVID-19, followed by recovery.
- Seasonality indicates regular sales changes throughout the year.



# Seasonal Analysis in Retail Sales

## Online Trends:

- Monthly online sales show peaks likely due to promotional events.
- Long-term trend shows consistent growth in online sales.
- Seasonal cycles reflect the impact of holidays and sales events.



# T test

- Conduct a t-test on the online monthly sales and offline monthly sales.
- The variances of the two sets of data are considered to be equal.
- There is a significant difference in the means of the two data samples.

Table: Statistical Test Results

Test	Value
Levene Statistics	0.6410
p-value (Levene)	0.4250
t Statistics	-41.4779
p-value (t-test)	1.066

# Regression analysis

- Statistically significant relationship.
- R-squared value of 0.108, indicating that online sales only modestly predict in-store sales.
- Contrary to expectations that online sales growth would weaken offline sales, both seem to be rising, possibly due to overall economic recovery post-pandemic.

```
=====
                        OLS Regression Results
=====
Dep. Variable:      Accumulated Sales      R-squared:                0.108
Model:              OLS                   Adj. R-squared:           0.092
Method:             Least Squares          F-statistic:             6.669
Date:               Mon, 04 Dec 2023        Prob (F-statistic):       0.0125
Time:               18:06:07               Log-Likelihood:          -525.45
No. Observations:   57                   AIC:                    1055.
Df Residuals:       55                   BIC:                    1059.
Df Model:            1
Covariance Type:    nonrobust
=====
                        coef      std err          t      P>|t|      [0.025      0.975]
-----
const                2.295e+04    1510.843     15.192     0.000     1.99e+04     2.6e+04
Accumulated Sales     0.4333         0.168       2.582     0.013     0.097     0.770
=====
Omnibus:              5.161    Durbin-Watson:           0.967
Prob(Omnibus):        0.076    Jarque-Bera (JB):         4.398
Skew:                 -0.669    Prob(JB):                 0.111
Kurtosis:              3.253    Cond. No.                  4.14e+04
=====

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 4.14e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Figure: Regression analysis

# Deseasonalized regression

- R-squared value of 0.162, indicating a stronger predictive power of online sales on in-store sales.
- Suggesting a meaningful relationship between online and deseasonalized in-store sales.

OLS Regression Results						
Dep. Variable:	y	R-squared:	0.162			
Model:	OLS	Adj. R-squared:	0.142			
Method:	Least Squares	F-statistic:	8.292			
Date:	Mon, 04 Dec 2023	Prob (F-statistic):	0.00618			
Time:	18:12:17	Log-Likelihood:	-404.17			
No. Observations:	45	AIC:	812.3			
Df Residuals:	43	BIC:	815.9			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	1.966e+04	2436.490	8.070	0.000	1.47e+04	2.46e+04
0	0.7825	0.272	2.880	0.006	0.235	1.331
Omnibus:	7.690	Durbin-Watson:	0.538			
Prob(Omnibus):	0.021	Jarque-Bera (JB):	7.036			
Skew:	-0.955	Prob(JB):	0.0297			
Kurtosis:	3.320	Cond. No.	7.44e+04			
Notes:						
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified						
[2] The condition number is large, 7.44e+04. This might indicate that there are strong multicollinearity or other numerical problems.						

Figure: Deseasonalized regression analysis

# Conclusion

- **Data-Driven Approach:** Utilized data from China's National Bureau of Statistics, analyzed via Python-based web scraping, covering the period from 2019 to 2023.
  - **Consumer Shift to Digital:** Identified a significant consumer preference shift towards e-commerce platforms.
  - **Impact of COVID-19:** Observed an acceleration in e-commerce adoption due to the pandemic, indicating lasting changes in consumer behavior.
  - **Seasonal Influence:** Detected the growing importance of seasonality in affecting both online and offline retail sales trends.
  - **Future of Retail:** Contributed insights crucial for understanding and anticipating the future dynamics of retail in a digital-dominated era.
- Impact of COVID-19: Observed an acceleration in e-commerce adoption due to the pandemic, indicating lasting changes in consumer behavior.